

Correcting the estimator for the mean vectors in a multivariate errors-in-variables regression model

Johannes Lutzeyer and Edward A. K. Cohen

Imperial College London

Abstract

The multivariate errors-in-variables regression model is applicable when both dependent and independent variables in a multivariate regression are subject to measurement errors. In such a scenario it is long established that the traditional least squares approach to estimating the model parameters is biased and inconsistent. The generalized least squares, ordinary least squares and maximum likelihood estimators (under the assumption of Gaussian errors) were derived in the seminal paper of [Gleser, 1981]. However, the ordinary least squares and maximum likelihood estimators for the mean vectors were incorrectly derived. In this short paper we amend this error, presenting the correct estimators of the mean vectors.

1 Introduction

[Gleser, 1981] presented the multivariate errors-in-variables regression model. We recap the formulation here and borrow the notation.

In a multivariate errors-in-variables model there are n observed random vectors $\mathbf{x}_i = (\mathbf{x}'_{1i}, \mathbf{x}'_{2i})'$, where \mathbf{x}_{1i} is $p \times 1$ and \mathbf{x}_{2i} is $r \times 1$, $i = 1, 2, \dots, n$. The assumed model is

$$\mathbf{x}_i = \begin{pmatrix} \mathbf{x}_{1i} \\ \mathbf{x}_{2i} \end{pmatrix} = \begin{pmatrix} \mathbf{u}_{1i} \\ \mathbf{u}_{2i} \end{pmatrix} + \begin{pmatrix} \mathbf{e}_{1i} \\ \mathbf{e}_{2i} \end{pmatrix} = \mathbf{u}_i + \mathbf{e}_i$$

$$\mathbf{u}_{2i} = \alpha + B\mathbf{u}_{1i} \quad i = 1, 2, \dots, n.$$

Here, the $(p+r)$ -dimensional random vectors \mathbf{e}_i , $i = 1, 2, \dots, n$ are i.i.d. with mean $\mathbf{0}$ and covariance matrix Σ_e . The vectors \mathbf{u}_{1i} , $i = 1, 2, \dots, n$, and \mathbf{u}_{2i} , $i = 1, 2, \dots, n$ are referred to as the *mean vectors*. The transformation parameters are the $r \times p$ matrix B and r -dimensional vector α . There are two distinct cases. The *no-intercept* model assumes α is known to be $\mathbf{0}$ and B and \mathbf{u}_{1i} , $i = 1, 2, \dots, n$ are unknown and to be estimated. The more general *intercept* model has α , B and \mathbf{u}_{1i} , $i = 1, 2, \dots, n$ as unknown and to be estimated.

As in [Gleser, 1981], notation is conveniently condensed by letting X_1 be the $p \times n$ matrix with columns \mathbf{x}_{1i} , $i = 1, 2, \dots, n$, letting X_2 be the $r \times n$ matrix with columns \mathbf{x}_{2i} , $i = 1, 2, \dots, n$, letting U_1 be the $p \times n$ matrix with columns \mathbf{u}_{1i} , $i = 1, 2, \dots, n$, and letting U_2 be the $r \times n$ matrix with columns \mathbf{u}_{2i} , $i = 1, 2, \dots, n$. We also let E be the $(p+r) \times n$ random matrix whose columns are \mathbf{e}_i , $i = 1, 2, \dots, n$. By constructing $X = (X'_1, X'_2)'$ and $U = (U'_1, U'_2)'$, we can represent the model as

$$X = U + E = \begin{pmatrix} \mathbf{0} \\ \alpha \mathbf{1}'_n \end{pmatrix} + \begin{pmatrix} I_p \\ B \end{pmatrix} U_1 + E,$$

where $\mathbf{1}_n$ is an n -dimensional column vector of ones. [Gleser, 1981] makes the following assumption:

the columns of E are i.i.d. with common mean vector $\mathbf{0}$ and common covariance matrix $\Sigma_e = \sigma^2 I_{p+r}$, where scalar $\sigma^2 > 0$ is unknown,

although Section 5 of [Gleser, 1981] demonstrates how this can be relaxed to a more general assumption of $\Sigma_e = \sigma^2 \Sigma_0$, Σ_0 known.

[Gleser, 1981] presents three approaches to estimating the unknown parameters U_1 , α , B and σ^2 of the multivariate “errors-in-variables” model. Firstly, in the *maximum likelihood* approach the columns of E are assumed i.i.d. multivariate normal allowing estimation of all four parameters. This estimation procedure is referred to as MLE. Secondly, in the *ordinary least squares* approach U_1 , α and B are chosen to minimize an orthogonally invariant norm of the residual matrix

$$R(\alpha, B, U_1; X) = X - \begin{pmatrix} \mathbf{0} \\ \alpha \mathbf{1}'_n \end{pmatrix} - \begin{pmatrix} I_p \\ B \end{pmatrix} U_1.$$

This estimation procedure is referred to as OLSE. Thirdly, in the *generalized least squares* approach α and B are chosen to minimize an orthogonally invariant norm of the normalized residual matrix

$$Q(\alpha, B; X) = (I_r + BB')^{-1/2} (X_2 - \alpha \mathbf{1}'_n - BX_1),$$

where $(I_r + BB')^{1/2}$ is any square root of $I_r + BB'$. This estimation procedure is referred to as GLSE.

The benefit of GLSE and OLSE approaches is they make no assumption on the distribution of the columns of E beyond those given for their mean vector and covariance matrix.

2 Correction

In this section it becomes necessary to define the following terms in the same way as Section 2 of [Gleser, 1981]. Let

$$\begin{aligned} C_n &= I_n, & \text{for the no-intercept model,} \\ &= I_n - n^{-1} \mathbf{1}_n \mathbf{1}'_n, & \text{for the intercept model,} \end{aligned}$$

let

$$W = XC_nX'$$

and let GDG' be the eigen-decomposition of W such that D is a diagonal matrix of the ordered eigenvalues and $(p+r) \times (p+r)$ matrix G is orthogonal. We partition G as

$$\begin{pmatrix} G_{11} & G_{12} \\ G_{21} & G_{22} \end{pmatrix},$$

where G_{11} is $p \times p$.

[Gleser, 1981] presents the following estimators:

$$\hat{B} = G_{21}G_{11}^{-1}, \tag{1}$$

$$\begin{aligned} \hat{\alpha} &= \mathbf{0}, & \text{for the no-intercept model,} \\ &= (-\hat{B}, I_r)\bar{x}, & \text{for the intercept model,} \end{aligned} \tag{2}$$

$$\hat{U}_1(\hat{\alpha}, \hat{B}) = (G_{11}G'_{11} X_1 + G_{11}G'_{21} X_2) C_n. \quad (3)$$

where $\bar{x} = n^{-1}X\mathbf{1}_n$. In Theorem 2.1 in [Gleser, 1981] (1) and (2) are shown to be the OLSEs and GLSEs for B and α , and (3) the OLSE for the mean vectors. For the no-intercept model this is correct. However, for the intercept model, while \hat{B} and $\hat{\alpha}$ are correct and it is correctly identified that the OLSE of the matrix of mean vectors U_1 is given by

$$\hat{U}_1(\hat{\alpha}, \hat{B}) = \left(I_p + \hat{B}'\hat{B}\right)^{-1} \left(I_p, \hat{B}'\right) \left(X - \begin{pmatrix} \mathbf{0} \\ \hat{\alpha}\mathbf{1}'_n \end{pmatrix}\right), \quad (4)$$

it is then wrongly stated that this is equal to the key expression given in (3).

In the following the term in (4) is manipulated to show it does not lead to the estimator in (3). Expanding (4) we get

$$\begin{aligned} \hat{U}_1(\hat{\alpha}, \hat{B}) &= \left(I_p + \hat{B}'\hat{B}\right)^{-1} \left(X_1 + \hat{B}'(X_2 - n^{-1}(-\hat{B}X_1 + X_2) \mathbf{1}_n\mathbf{1}'_n)\right) \\ &= \left(I_p + \hat{B}'\hat{B}\right)^{-1} \times \\ &\quad \left(X_1 + n^{-1}\hat{B}'\hat{B} X_1 \mathbf{1}_n\mathbf{1}'_n + \hat{B}'X_2 - n^{-1}\hat{B}'X_2 \mathbf{1}_n\mathbf{1}'_n\right). \end{aligned} \quad (5)$$

Since G is orthogonal, considering the top left block of the equality $G'G = I_{p+r}$ we get the relation $G'_{11}G_{11} + G'_{21}G_{21} = I_p$. Using this and $(G_{11}^{-1})' = (G'_{11})^{-1}$ the following simplification arises:

$$\begin{aligned} \hat{B}'\hat{B} &= (G_{11}^{-1})' G'_{21}G_{21}G_{11}^{-1} = (G_{11}^{-1})' (I_p - G'_{11}G_{11}) G_{11}^{-1} \\ &= (G_{11}^{-1})' G_{11}^{-1} - I_p. \end{aligned} \quad (6)$$

From (6) we have (5) simplify to:

$$\begin{aligned} \hat{U}_1(\hat{\alpha}, \hat{B}) &= G_{11}G'_{11} \left(X_1 + n^{-1} \left((G_{11}^{-1})' G_{11}^{-1} - I_p\right) X_1 \mathbf{1}_n\mathbf{1}'_n \right. \\ &\quad \left. + (G'_{11})^{-1} G'_{21}X_2 - n^{-1} (G'_{11})^{-1} G'_{21}X_2 \mathbf{1}_n\mathbf{1}'_n\right), \end{aligned}$$

giving the final result

$$\hat{U}_1(\hat{\alpha}, \hat{B}) = n^{-1}X_1\mathbf{1}_n\mathbf{1}'_n + (G_{11}G'_{11} X_1 + G_{11}G'_{21} X_2) C_n. \quad (7)$$

We are therefore left with the additional term $n^{-1}X_1\mathbf{1}_n\mathbf{1}'_n$, which is not present in (3). Using the correct expression (7), Theorem 2.3 in [Gleser, 1981], which states this to also be the MLE, is now also correct.

References

[Gleser, 1981] Gleser, L. J. (1981). Estimation in a multivariate “errors in variables” regression model: large sample results. *The Annals of Statistics*, 9:24 – 44.